ICIC Express Letters
Volume **11**, Number **4**, April **2017**

ICIC International ©2017 ISSN 1881-803X
pp. 899–904

LATENT SEMANTIC ANALYSIS OF RESEARCH PAPERS ON SMART FACTORY

Tai-Woo Chang* and Hyun-Lim Yang

Department of Industrial & Management Engineering
Kyonggi University
154-42, Gwanggyosan-ro, Yeongtong-gu, Suwon, Gyeonggi 16227, Korea
*Corresponding author: keenbee@kgu.ac.kr

Received December 2016; accepted February 2017

Abstract. *As the growing competition of manufacturing businesses, many countries are trying to adopt smart factory as a way of breakthrough. In Korea, government announced 8 foundation techniques of smart manufacturing, and made policy to encourage companies to adopt them. Although there have been many researches about smart factory and smart manufacturing, there are little basic statistical research. In this study, we analyzed 371 published Korean research papers about smart factory using latent semantic analysis (LSA), and found the research trend of this field according to five topics which are categorized by LSA methodology.*

Keywords: Smart factory, Trend analysis, Text mining, Topic model, LSA

1. **Introduction and Related Works.** Many countries are building smart factories through converging information and communication technologies (ICT) to strengthen the competitiveness of the manufacturing industry. Many strategies such as 'Industry 4.0' of Germany and related researches enable the construction of smart factories. Various promotion policies have been presented in South Korea as well, with the 'Manufacturing Innovation 3.0' being the lead. This phenomenon attracts attention not only from the national level but also from academia, and related active researches are ongoing.

Since smart factories are fused in various fields, there are diverse viewpoints and research fields for researching them, and it is difficult to grasp the flow of these industries and the direction of research at the same time. In addition, appropriate statistical data for objectively diagnosing and properly interpreting these are also absent [1].

Therefore, in this study, we would like to analyze the keywords that appeared in Korean domestic journals related to smart factory, grasp the research trend up to the present, and identify the main research fields by time. In addition, by analyzing the academic papers which are authorized intellectual achievement, we intend to grasp current research situation through objective statistical processing, and to lay the foundation for future research direction of smart factory. To do this, we used latent semantic analysis (LSA) as an analytical method to infer some topics about smart factory and identify related trends.

Academic papers are kinds of the intellectual structures. They are systematic description of the research results conducted by researchers and published in academic journals after peer reviews. In the past, it was very difficult to collect and analyze a large number of academic papers collectively. However, recently, unstructured data explosively increased, and many text mining techniques have been developed and analyzed.

Cho and Kim [2] analyzed the correlation between industrial engineering related keywords of papers in a journal, IIE Transactions, and performed clustering the keywords by using text mining technique. However, in the study, there was a limitation to analyze the trend of the relevant research field in general by analyzing a specific journal as a target. As a similar previous research, Kim and Jang [3] analyzed the research trend in Korean industrial and management engineering fields by using LDA (latent Dirichlet allocation) methods. Also, there was an attempt to identify research trends in the field of operation management using LSA method [4].

This paper is organized as follows. Section 2 introduces the research methods and procedure. Section 3 explains the results of trend analysis. Concluding remarks are given in Section 4.

2. **Research Methods.** In this study, we used the LSA method to find some topics and to analyze the research trends of smart factory. First of all, pre-processing steps such as removal of redundant papers, elimination of unnecessary terms, standardization of terms and term-document matrix (TDM) generation are essential for collected papers. In addition, weight transformation for each analysis technique is needed. After performing LSA and validation, we visualized and analyzed the results.

2.1. **Data collection.** To analyze the research trends related to smart factory, we collected abstracts and information of papers through two portal sites of academic information, RISS and NDSL in Korea. We collected papers, title and abstract of which have Korean keywords including 'smart factory', 'smart manufacturing', 'smart logistics', 'Manufacturing Innovation 3.0', 'Industry 4.0'. There were a total of 558 papers found. After duplicate papers and papers not related to manufacturing and ICT industry were removed, 371 Korean papers were analyzed.

2.2. **Data preprocessing.** Based on the abstract information of collected papers, preprocessing of data is necessary to execute text mining. First, duplication of these data is eliminated, and the key value is given for each paper for later analysis. Subsequently, the abstract information of the paper and the other information are separated, and the dictionary to be used for morpheme analysis is called up. Based on the results of morpheme analysis, we create TDM by creating corpus, which is a collection of text data that the computer understands. In addition, it is necessary to delete the unnecessary terms and standardize remaining terms in this process.

When generating the TDM, there are various methods for assigning weights to the occurrence frequencies of terms for each document. In this study, we assigned the weight of TF-IDF method multiplied by term frequency (TF) and inverse document frequency (IDF) when executing LSA.

2.3. **Latent semantic analysis (LSA).** LSA is designed to improve initial information retrieval and query retrieval performance by detection of relevant documents based on terms found inqueries [5]. It uses the singular-value decomposition (SVD) technique, reduces the dimensions and looks for potential meaning within the document.

LSA is a theory and method for extracting and representing the contextual-usage meaning of words by statistical computations applied to a large corpus of text [7]. The underlying idea is that the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other [6]. LSA assumes that words that are close in meaning will occur in similar pieces of text.

Typically, LSA has been used in a couple of ways [6]: to find segments from large text data by grouping the terms which have similar contextual usage, or as a model of statistical process and representation of underlying knowledge.

LSA is a multivariate statistical analysis method, and analyzes the document vectors on semantic space while maintaining the explanatory power of data. By using SVD technique, we can redefine the semantic space by ignoring some singular values which have low influence on it. That is, SVD is used to reduce the number of rows (columns) while preserving the similarity structure among columns (rows). When projected into a reduced space, a document or word with a similar co-occurrence pattern approaches the nearest location. So, we can easily find the patterns and related contents from the documents on the new space. With selected singular value and redefined dimension, it is possible to explore the noise-cancelled 'latent semantics' or topics.

In this study, we explored five topics through LSA based on the abstract data of Korean research papers about smart factory, and tracked the research trends and changes on smart factory by year.

3. **Research Results.** To analyze the research trend of smart factory, we compared the key terms of papers by 2010 and the ones after 2010 year by year. 2011 is the year when 'Industry 4.0' originates by Germany to promote the computerization of manufacturing. Papers which are related to smart factory have been published year by year as shown in Table 1. It shows a moderate upward trend.

Table 1. Number of papers each year related to smart factory

∼2010	*2011*	*2012*	*2013*	*2014*	*2015*	*2016 (∼Sep.)*
81	30	41	44	50	74	51

LSA uses a statistical method to probabilistically infer the associations between hidden words using the words that exist in the document. By doing this, it is possible to create new axes that can maintain good association between words and to analyze and output the results by rearranging the documents on the newly generated vector space. All terms in this study are categorized into three to ten axes of each dimension that are newly defined by LSA. We reviewed the terms of all dimensions and selected five-dimensions (in Table 2) by binding the meaning of terms contained in each axis to the same topics.

Table 2. Ten major terms of five axes

No	*Major terms and the score*									
1	Manufacturing industry	Korea	Germany	USA	policy	China	investment	industry	competitiveness	Japan
	0.2724	0.2220	0.2159	0.1967	0.1981	0.1766	0.1713	0.1679	0.1621	0.1518
2	Ratio	Finance	inventory	level	labor	economy	growth	import	deepening	increase
	0.2423	0.2321	0.2217	0.1772	0.1719	0.1707	0.1642	0.1421	0.1420	0.1415
3	bio-engineering	technology transfer	government funded institute	scientific	table	knowledge base	university	commercialization	best practice	venture business
	0.3404	0.2647	0.2467	0.2438	0.2287	0.2001	0.1913	0.1872	0.1872	0.1667
4	mobile	smart	system	service	design	RFID	device	sensor	IoT	Internet
	0.2346	0.2177	0.2175	0.2170	0.1988	0.1863	0.1728	0.1690	0.1574	0.1455
5	first	innovation	conglomerate	service industry	SME	enterprise	difference	importance	survey	new product
	0.3719	0.2752	0.2638	0.2562	0.2473	0.1936	0.1861	0.1827	0.1813	0.1803

We defined five potential topics that each axis implies. They are 'policies by country', 'economic change', 'research and development', 'information technology', and 'enterprise innovation', and the research trends were examined based on these. Table 3 shows the relationship of highly relevant terms probabilistically inferred by LSA.

Because LSA infers the topics of documents based only on the key terms, in case of long document it can have high values in all dimensions in the vector space, regardless of the dimension value. To compensate for these distortions, we modified the values of TDM via log (TF) and the weights for the number of words repeated in a document before performing the LSA. In addition, we modified the weights for the key terms, which appear frequently in each paper, such as 'smart', 'manufacturing', and 'research' by using

Table 3. Major keywords and inferred potential topics

Major keywords	***Topics***
manufacturing industry, Korea, Germany, USA, policy, China, investment, industry, competitiveness, Japan	Policies by Country
ratio, finance, inventory, level, labor, economy, growth, import, deepening, increase	Economic Change
bio-engineering, technology transfer, government funded institute, scientific, table, knowledge base, university, commercialization, best practice, venture business	R&D
mobile, smart, system, service, design, RFID, device, sensor, IoT, Internet	Information Technology
first, innovation, conglomerate, service industry, SME, enterprise, difference, importance, survey, new product	Enterprise Innovation

the value of IDF. The titles of representative papers with high values for each dimension analyzed through this are shown in Table 4.

Based on the five axes and topics defined before, we analyzed year-on-year research trends on the topics as shown in Figure 1. The x-axis represents the yearly flow of time and the y-axis means the ratio at which papers related to the topic are published in the year. In order to reduce the noise, we utilized only the top 70% papers of the absolute values among the ones with positive values in the dimension, and added them to the ratio estimate. Each paper could appear in duplicate at all dimensions. To facilitate the analysis of the results, trend lines for each topic are shown using the LOESS (locally estimated scatter-plot smoothing) algorithm, one of the non-parametric regression analysis algorithms.

As shown in Figure 1, research on information technology related topic has been actively researched recently, and it can be seen that the related research is continuously increasing. Prior to 2000, only some of the studies related to smart factory were covered, but it has been consistently high percentage since the beginning of the 2000s and continues to the present. In addition, the frequency of publication of papers on the topic related to policies by country is also increasing. This can be interpreted as a result of the analysis of cases studies of the United States and Germany, which are known as the leaders of the smart factories and the research on case studies of China called the factories in the world being actively conducted.

Other topics show a decreasing trend. In particular, topic related to enterprise innovation has been actively studied to the greatest extent in the early 2000s, but the proportion of related research has decreased sharply since 2006. Topics related to economic change and R&D have also been actively studied in the past, but the proportion of research has been steadily decreasing over the last five years.

4. **Conclusions.** Smart factories have recently been explored in many ways as an innovation tool in the manufacturing industry, and many companies are working to build it. However, smart factory is a fusion of many technologies, so there are various research directions and it is difficult to grasp the trends at a glance. In this study, we collected research papers in the field of smart factory and analyzed them through text mining to analyze and forecast research trends. Especially through the LSA, we set five research topics in the smart factory field and analyzed trends on these topics.

As a result, research on topics such as enterprise innovation, economic change, and R&D has been steadily decreasing over the past five years, and research on topics of information technology and policies by country has been steadily increasing. We can see that introduction of information and communication technologies at smart factory and bench marking of overseas best practices have been actively underway recently.

Table 4. Titles of related papers and inferred topics of five dimensions

Topics	***Titles of papers***
Policies by Country	- Growth Factor Analysis on the Korean Manufacturing Industry - Progress and Implications of American Renaissance of Manufacturing Industry - Critical Factors of Competitiveness of the German Manufacturing Industry - Standardization Trends and System Architecture of Smart Factory - Analysis of ICT Startup Environments in Major Chinese Areas
Economic Change	- International Comparative Analysis of Korean Industry Competitiveness - Status Evaluation and Implications of Manufacturing Innovation Policies - An Analysis of Logistics Cost Structure by Iceberg in Korean Firms - International Comparative Analysis of Trend of High Value-Added Manufacturing and Export Competitiveness - Recent Trends of Protectionism and Korea's Countermeasures
R&D	- Actual Utilization of Intellectual Properties in Korea and Its Implications - Enhanced Performance of International Competitiveness and Complementary Points of the Parts Industry - Development and Design of a Vision System for Inner Defect Inspection of Paper-Cup-Manufacturing Smart Factory - Survey for R&D Activities of Foreign Firms in Korea - Development of On-Line Monitoring System Using Smart Material
Information Technology	- A Study on ICT-Based Smart Farm Factory Integration Platform - An Exploratory Study on Application and Development Direction of Smart Logistics in Autonomous Vehicle Technology - Smart Work Based System Architecting of Shipbuilding Manufacturing Execution System - Development Smart Sensor & Estimation Method to Recognize Materials - A Study on Realization of System in Wireless Location Awareness Technology Using Ubiquitous Active RFID
Enterprise Innovation	- Survey on Technological Innovations in Korea - A Study on Effect of Product Design and a Primary Factor of Quality Competitiveness - Enhanced Performance of International Competitiveness and Complementary Points of the Parts Industry - Status Evaluation and Implications of Manufacturing Innovation Policies - Integrated Operation Technology Development of Smart Factory Application Connected with Manufacturing Big-data

The result of this study has significance in that it was possible to read clear trends by more numerically and statistically analyzing the existing research papers related to smart factory. In addition, by analyzing potential topics in the papers, we were able to grasp research trends related to smart factory more diversely. And the statistical framework for predicting more reliable research directions has been prepared by combining these.

The number of research papers related to smart factories in Korea was still insufficient and trend analysis had limitations. Also, with the mixture of terms, we could not utilize

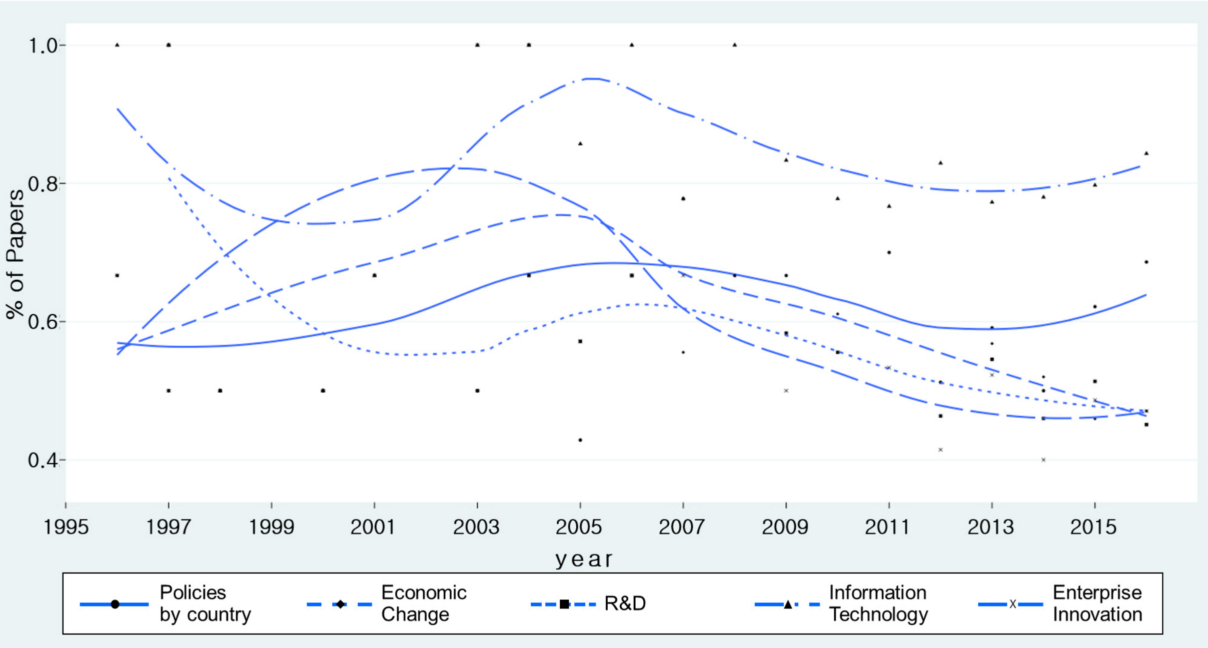

FIGURE 1. Changes in the ratio of research papers on topics by year

all smart factory related papers for retrieval and analysis. The fact that the abstract data is inconsistent and the amount of information is different is also a limitation of the study. If the data which can be analyzed in the future increase significantly, we expect to be able to analyze research trends using advanced analytical techniques that can be effectively applied to large data such as LDA and Autoencoding techniques.

Acknowledgement. This work was supported by Kyonggi University Research Grant 2016.

REFERENCES

[1] J. Lee and J. Park, Future direction of classification on smart factory related industry, *Proc. on 2016 Spring Conference of Korean Institute of Industrial Engineers*, pp.1493-1515, 2016.
[2] S. Cho and S. Kim, Finding meaningful pattern of key words in IIE transactions using text mining, *Journal of the Korean Institute of Industrial Engineers*, vol.38, no.1, pp.67-73, 2012.
[3] S. K. Kim and S. Y. Jang, A study on the research trend in domestic industrial and management engineering using topic modeling, *Journal of the Korean Management Engineering Society*, vol.21, no.3, pp.71-95, 2016.
[4] S. S. Kulkarni, U. M. Apte and N. E. Evangelopoulos, The use of latent semantic analysis in operation management research, *Decision Science*, vol.45, no.5, pp.971-994, 2014.
[5] S. Deerwester, S. Dumais, G. Furanas, T. Landauer and R. Harshman, Indexing by latent semantic analysis, *Journal of the American Society for Information Science*, vol.41, no.6, pp.391-407, 1990.
[6] T. K. Landauer, P. W. Foltz and D. Laham, Introduction to latent semantic analysis, *Discourse Processes*, vol.25, pp.259-284, 1998.
[7] T. K. Landauer and S. T. Dumais, A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge, *Psychological Review*, vol.104, pp.211-240, 1997.